COMPUTER AIDED COGNITION TO SUPPORT PROBLEM-CENTERED DECOMPOSITION OF COMPLEX PROBLEMS

Michael McNeese
and
David L. Hall
School of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802

Complex problems such as analysis of military situation assessment, homeland defense, diagnosis of the health of complex systems, medical diagnosis, and environmental monitoring require the ability to utilize a wide variety of data such as signals, images, textual information, and scalar data. The rapid evolution of micro-scale sensors, wideband communications, and microprocessors enables the collection and dissemination of huge amounts of data to be provided to a human analyst. Unfortunately, the analyst cannot directly understand nor process the data. Instead, analysts reason about high-level abstractions via language. A challenge exists to decompose general problems into detailed models that link to specific types of data (viz., problem centered decomposition) and to compose data into meaningful relationships to assist the understanding of semantic representations of abstract concepts. This paper discusses the challenge of problem-centered analysis (including problem centered decomposition and problem centered composition) and describes our efforts to develop cognitive aids to assist the analysis process for improved understanding of complex problems.

INTRODUCTION

A fundamental paradox exists in information fusion. Information fusion in this context may be used in traditional areas such as national defense, counter-intelligence, situation assessment for tactical military applications (Hall and Llinas 2001), or non-Department of Defense (DoD) applications such as environmental monitoring, technology assessment for business applications or related areas (Llinas and Hall 1994). The paradox is that information analysts are drowning in a sea of data but unable to obtain the knowledge that they need to address difficult problems. This has often be referred to as the data overload dilemma (Kuperman, 2001) or more recently framed "cogmenutia fragmentosa" (McNeese & Vidulich, 2002).

On one hand an unprecedented capability exists to collect data via distributed sensors, commercial information providers (e.g., AccuWeather, Library Services, commercial search businesses), human sources, or Internet resources. Smart micro-scale sensors (Jones 1995), wireless communications, and global Internet accessible resources enable the entire earth to be a potential information resource (the I-earth observatory). Such information is available literally at the fingertips of the analysts. In particular, the Internet has exceeded one billion web pages, with a continuing exponential increase. Analysts are literally swamped with data. They have a wide variety of choices to make as to what is *useful* and *usable,* given the context of what they are trying to *understand* (Woods, 1998).

On the other hand, the glut of data can be overwhelming and may inadvertently promote poor decision processes (Ferran 1999). Studies of decision-making under stress have shown that too much information can cause ineffective decision styles. An example is the *hyper-vigilance mode*, in which a decision-maker frantically searches for new information, without taking time for reflection and thoughtful analysis of existing data. The huge glut of rapidly changing data via the Internet may encourage this type of response. Alternatively, a decision-maker may feel overwhelmed with new information and simply ignore new data. Thus, in a rich atmosphere of data, decision-makers are suffocating for knowledge (McNeese and Vidulich 2002). They may have a large amount of cognitive readiness available to fuse multiple information sources but in fact their meta-cognition (McNeese, 2000) may be very limited. This often makes predictions about "what to do next" daunting.

Through the use of contemporary cognitive systems engineering approaches (e.g., *The Living Lab Framework,* McNeese, 2002), a vision for information-based fusion is to transform the current analyst dilemma to an ideal situation in which the analyst can directly access information. We seek to develop tools, models and techniques to allow an analyst to effectively use the entire earth's observing resources for situation assessment.

TRADITIONAL CONCEPT OF INFORMATION-BASED FUSION

In order to develop tools and techniques for effective assimilation of data, it is instructive to reconsider the traditional process of information fusion (Hall 1992). Traditionally, data fusion has been described as a hierarchical inference process involving a transformation of energy (e.g., observed by active and passive sensors) to information or knowledge for human use. Figure 1 depicts this transformation. Similarly, Figure 2 depicts the standard Joint Directors of Laboratories (JDL) data fusion process model (Kessler 1992). By implication data fusion involves the ingestion of data (shown on the left hand side of Figure 2), processing by various "levels" of data fusion functions, and presentation of this information for utilization by analysts (shown on the right hand side of Figure 2).

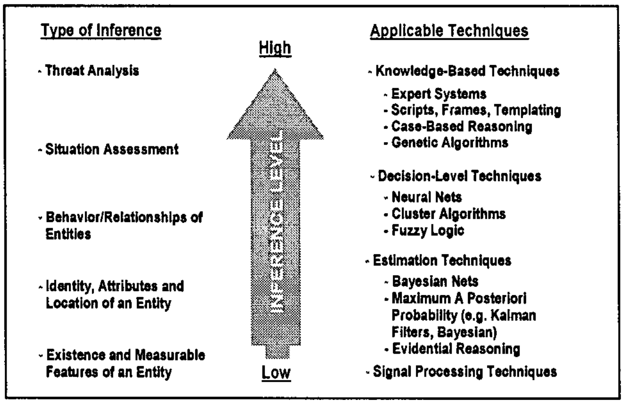

Figure 1: Inference Hierarchy

Recently, several user-centric attempts have sought to evolve the data fusion process to one that is focussed on the analyst/decision-maker, and link user needs for information to sensor tasking. Hall, Hall & Oue (2000) argued the need for more explicit consideration of the human side of information fusion and they introduced the concept of level 5 processing in the JDL model to explicitly support the human user. Hall & Garga (1999) described new perspectives on level 4 processing and suggested the need for improved linkage between human information needs and sensor tasking and algorithm control. DARPA's on-going dynamical tactical targeting (DTT) program has a component to develop new algorithms for improving the link between information needs and sensor tasking. Finally, the TRIP model (Fabian and Eveleigh 2001) has been developed to show stronger links between information needs and sensor tasking.

HUMAN-CENTERED INFORMATION FUSION

Information fusion requires the formations of what we have referred to as strong mental models on the behalf of the information analyst. A strong mental model utilizes the strength of both perceptual recognition and language-articulation processes to construct meaning from impinging real world events. In contrast, a weak mental model consists of only one of these dimensions but not both. When computer interfaces are designed for complex information fusion *applications, it is often the case that their design is not human-*centered. If human-computer interaction is a factor in design, it is typically considered only from the weak mental model perspective. In turn this produces tools, support systems, or interfaces that are less than desirable, falling far short of providing comprehensive and effective assistance for the information analyst. Although perceptual recognition is extremely important (McNeese, 2000), this paper specifically highlights the role of semantic, language, and explanation-based aspects of information fusion.

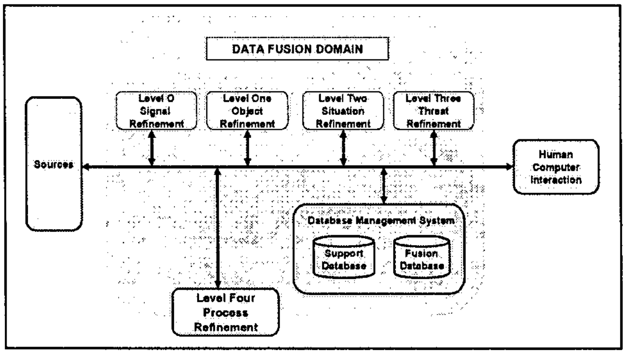

Figure 2: JDL Data Fusion Model

In order to improve human-centered information processing, we introduce here the concept of semantic-based information fusion (shown in Figure 3). The analyst is considered to be the center of an on-going emergent and evolutionary process that accesses enormous amounts (peta-bytes) of collected data to recognize a problem of interest.

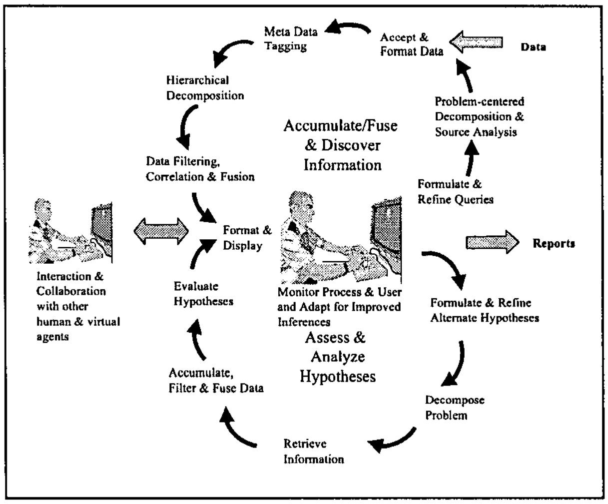

Figure 3: Human-Centered Semantic Information Fusion

Given the level of time pressure in a situation, an analyst may quickly utilize recognition-based models (e.g., Zsambok & Klein, 1997) to develop information fusion. Alternatively, if time allows an analyst may engage higher level cognition to construct (and trv to validate) hypotheses regarding an evolving situation or threat. This process is described as semantic information fusion because the analyst performs analysis using in-depth, higher order cognition that require construction of semantic concepts (viz., words, phrases, and scenarios). Thus, while data such as signals, images, and text play a major role in the recognition-analysis process (e.g., recognition-primed decision models, Zsambok & Klein, 1997) the interpretations by the analyst are defined in terms of language, logic, or explanations, via joint composition of articulated meanings with other (often remotely located) team members. This has been referred to by others as distributed cognition (Hutchins, 1995). The following discussion (and diagram in Figure 3) shows a single analyst at the center of this process. However it is recognized that a team of analysts may construct this process, and that in fact these analysts may not be physically co-located.

Accumulate/Fuse and Discover Information

The top part of the processing cycle shown in Figure 3 involves the collection, accumulation, and fusion of data with semi-automated (human-aided) discovery of knowledge. This process may be part of a background information collection, accumulation, and discovery process performed over long periods of time. The result of this process may be the "population" of a data warehouse aimed at providing background and real-time information to support analysis. The process is shown as a cyclical process that involves seven steps. Note that these steps are illustrated as being sequential. However, in an actual system, the steps could be integrated. In addition, these functions may be automatically performed by processing functions (e.g., intelligent agents), by human analysts, or by a hybrid human/computer analysis process.

Accept and Format Data – Incoming and collected data must be accepted and formatted by the distributed collection/processing system. Here it is assumed that data includes data (signals, images, vectors, and scalar data) observed by sensors, sensor reports (e.g., processed data), reports and information submitted by human analysts, information gathered from open-source information, and models. Data formatting includes functions such as unit conversions, sensor model (or platform) corrections, bias removal, information reformatting or translations.

Meta Data Tagging and Transformations – Data may be tagged with explanatory information to augment the data. Examples include the use of a pre-specified ontology to assist in characterizing the information, extraction of key words and descriptors, annotation of the data with parametric information related to location, identity, or characteristics. Currently, these types of tagging are usually performed only on textual information. In future systems, advanced processing may be performed to automatically characterize images (e.g., to recognize objects and characterize background scenes). The development of data tags provides the basis for rapid data retrieval and correlation.

Hierarchical Decomposition and Pattern Recognition – Another step involves decomposition of the data into components. For image data this may entail image segmentation and hierarchical de-composition into smaller components. For textual data this may include development of a hierarchy of smaller units. Ideally, such a decomposition would relate to a hierarchy of concepts (from very abstract to very concrete, e.g., the abstraction hierarchy, Rasmussen, 1999).

Data Filtering, Correlation, and Fusion – Data may be filtered, correlated, and possibly fused. Examples of correlation include identifying pieces of information that relate to a specific area of interest, event, or situation. A component of this function may include link analysis to identify and link together related concepts or semantic "quanta" of information.

Format and Display – In order for a human to access information, it must be formatted for display and presentation. Commercial systems such as geographical information systems (GIS) provide an example of the types of functions that can be performed to present geographical and related information to a human. Rapid advances in web-based services provide an evolution of hypertext, graphics, icons, and similar techniques to improve the access of a human to data. However, as previously indicated, this is an under researched area. The call for a Level 5 process in data fusion identified a number of potential research areas to advance the link between humans and data.

Query Formulation and Refinement – The on-going data accumulation, fusion, and information discovery does not occur in a background. Ideally, the analyst provides guidance and feedback to the automated process. Formulating and refining general and specific queries (i.e. default queries and specific queries to support evolving investigations) can provide guidance to the data discovery process. In addition, the analyst can inform the system regarding preferences, changes to ontology, identification of new areas of interest or key words and rate the performance of the system in establishing patterns or linkages.

Problem-centered Decomposition and Source Analysis - The concept of problem-centered decomposition can be used to decompose general queries into specific queries, tasking for sensors, and tasking for information sources. Such decomposed types of information requests may guide the overall ingestion, accumulation, fusion and knowledge-discovery process.

Central to the on-going upper loop of accumulation and discovery is the role of analysts to assist in the interpretation of the information and tasking the accumulation and discovery process. Ideally, the actions of the analyst and the on-going process should be monitored with semi-automated adaptation to improve the effectiveness of the analysis process and the discovery process. This is shown in Figure 3 below the picture of the analyst. In addition, the analyst may interact with other agents or analysts. Hence, a domain specialist may interact with specialists who are knowledgeable about specific sensing systems or types of data. These external collaborators could conceptually be other human agents, software models, or intelligent software agents.

Assess and Analyze Hypotheses

The assessment and analysis process is shown in the lower part of Figure 3. Conceptually an analyst will be tasked (or self-tasked) to analyze an evolving situation or threat. The analyst has access to an enormous amount of information via the continuing data acquisition and warehousing process via the upper cycle shown in Figure 3. However, the analyst is not interested solely in acquiring massive amounts of data. Instead, the analyst seeks to develop hypotheses regarding an evolving situation and to assess and analyze these hypotheses.

The process for analyzing these hypotheses is illustrated in the cyclical process shown on the bottom of Figure 3. The process involves a number of steps summarized below. As with our previous discussion, these steps are not necessarily performed in sequence. In addition, it is understood that these steps may be performed by a human, an automated computing process, or by a hybrid human/computer operation. The steps are summarized below.

Formulate and Refine Alternative Hypotheses – The analyst continually formulates alternate hypotheses (or tentative explanations or interpretations) of the evolving situation. This formulation assists the analyst in focusing his attention on possible situations and assists in defining what information must be collected or analyzed.

Decompose Problem – The decomposition of the problem from a general query or hypothesis to specific sources of needed information is the inverse of the inference hierarchy shown in Figure 1. Problem-centered decomposition involves transformation from a general query to specific sensor or information source tasking.

Retrieve Information – Given specific essential elements of information required to address the query or hypothesis, the system needs to assist the analyst in the retrieval of this information.

Accumulate, Filter, and Fuse Data – In support of the evolving analysis, data and information are accumulated and filtered. Components of information fusion are required to sort, accumulate, correlate, and fuse information that may support or refute alternative hypotheses. In this case transformations are sought between low-level data and more general representations and inferences.

Evaluate Alternative Hypotheses – Alternative hypotheses may be evaluated in a quantitative way to provide support or refutation of the alternatives. Simulation tools may be used to project the hypotheses forward in time to determine the consequences and likelihood. Evidential reasoning and course of action analysis is required in this step.

Format and display – Finally, the inferences and supporting data must be formatted and displayed to the analyst for evaluation, refinement, and continuation of the process. When complete (or at least suitable for reporting), the analyst may create a report for evaluation by other analysts or decision-makers.

Here again, a team of analysts, supported by external human analysts, models, and software agents, may conduct this process. The analysis process should be monitored to determine the performance of the process (and the performance and preferences of the human analyst) to improve the short term and long term effectiveness of the analysis. Ideally, when collaborating with external resources, the analyst should not need to know (nor care) whether his collaborator is a human, a software model, or an intelligent software agent.

COGNITIVE AIDS TO ENABLEINFORMATION-BASED FUSION

Our current research is focused on understanding the cognitive processes involved in information fusion and development of computer-based cognitive aids that can support and improve the cognitive process. A summary of potential innovations for the information ingestion and hypothesis evaluation cycles is provided in Figure 4. Each of the steps in the dual data accumulation/analysis processes provides an opportunity for improvements and innovations.

Figure 4: Potential Innovations to Improve Problem-Centered Decomposition

Process Cycle	Process Component	Innovation Concepts and Techniques
Accumulate, fuse & discover information	Accept & Format Data	- Rapid format and ingestion methods - Standardization (e.g., via XML or equivalent)
	Meta Data Tagging & Transformations	- Tag data at a semantic level to allow context-based retrieval and associations - Associate memory mapping - Map images to semantic concepts via multi-resolution wavelets and semantic-to-bit classification map (Wang) - Dynamic ontology modification (F. Fonseca) - Efficient text preparation (L. Giles)
	Hierarchical Decomposition	- Link analysis (D. Davenport) - Model-based decomposition
	Data Filtering, Correlation & Fusion	- Multi-INT rapid correlation & fusion) - Hybrid reasoning for context-based fusion

Figure 4: Potential Innovations to Improve Problem-Centered Decomposition

Process Cycle	Process Component	Innovation Concepts and Techniques
Assess & Analyze Hypotheses	Format & Display	- Multi-sensory data representation - Deliberate synesthesia - Full-immersion 3-D representations - Human/Agent communication language - Collaboration tools - Gesture recognition/GIS communication - Visualization of non-physical concepts
	Formulate & Refine Queries	- Agent-based generation of alternative hypotheses - Logical templates for "fill in the blank" hypotheses
	Problem-Centered Decomposition & Source Analysis	- PCD methodology defined for intelligence analysis - PCD decomposition models based on systems engineering design concepts - Entity to observable mapping templates - Event logical templates (hybrid reasoning)
	Process Monitoring & Adaptation	- Affective computing - Analyst characterization & preference modeling - Measures of effectiveness (MOE) monitoring - Team performance modeling - Multi-objective optimization (e.g., goal programming)
	Formulate & Refine Alternate Hypotheses	- Agent-based generation of alternative hypotheses - Logical templates for "fill in the blank" hypotheses - Case-based reasoning for hypothesis definition - Biologically inspired ant swarm outlier exploration
	Problem-centered Decomposition & Source Analysis	- PCD methodology defined for intelligence analysis - Problem decomposition models (based on systems engineering design concepts) - Entity to observer mapping templates - Event logical templates (hybrid reasoning)
	Retrieve Information	- Niche search engines - Dynamic ontology generation & modification - Semantic language-to-image feature transformation & retrieval
	Accumulate & Filter Data	- Data fusion tools - Rapid multi-INT association, correlation & fusion - Hybrid reasoning pattern recognition
	Evaluate Alternative Hypotheses	- Simulation-based hypothesis prediction (consequence reasoning) - Curmudgeon agent evaluators - Negative reasoning emulators

ACKNOWLEDGEMENTS

This research was supported in part by a grant from the Lockheed Martin Corporation (Management and Data Systems).

REFERENCES

Fabian, B. and T. Eveleigh (2001). *Estimation and ISR process integration: A process model for information exchange between situation assessment and collection managment.* MSS National Symposium on Sensor and Data Fusion, San Diego, CA.

Ferran, B. (1999). U. S. g. National Reconnaissance Office. Chantilly, Va, Available on video tape from the NRO office in Chantilly, VA.

Hall, D. (1992). *Mathematical Techniques in Multisensor Data Fusion.* Norwood, MA, Artech House, Inc.

Hall, D. and A. Garga (1999). *New perspectives on level four processing in data fusion systems.* SPIE in AeroSense 1999, Orlando, FL.

Hall, D. and J. Llinas, Eds. (2001). *Handbook of Multisensor Data Fusion.* New York, CRC Press.

Hall, M. J., S. A. Hall, et al. (2000). *Removing the HCI bottleneck: How the human computer interface (HCI) affects the performance of data fusion systems.* 2000 MSS National Symposium on Sensor and Data Fusion, San Diego, CA.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

Jones, A. (1995). Micro-electrical-mechanical Systems (MEMS): A DoD Dual Use Technology Assessment. Washington, DC, Department of Defense.

Kessler, O. e. a. (1992). Functional description of the data fusion process. Warminster, PA, Office of Naval Technology, Naval Air Development Center.

Kuperman, G. G. (2001). Cognitive systems engineering for battlespace dominance. In M. D. McNeese & M. Vidulich (Eds.), *Cognitive systems engineering in military aviation environments: Avoiding Cogminutia Fragmentosa* (pp. 230-241). Wright-Patterson Air Force Base, OH: HSIAC Press.

Llinas, J. and D. Hall (1994). *Dual-use of data fusion technology: applying the JDL model to non-DoD applications.* Fourth Annual IEEE Dual-Use Technology and Applications Conference, SUNY Institute of Technology, Rome, NY.

McNeese, M. D. (2000). Socio-cognitive factors in the acquisition and transfer of knowledge. *Cognition, Technology, and Work, 2,* 164-177.

McNeese, M. D. (2001). *Situated cognition in distributed multi-national, multi-cultural teams: A research prospectus informing socio-cognitive technologies.* Final report submitted to the United States Navy, SPAWAR Center, San Diego, CA.

McNeese, M. D. (2002). Pursuing medical human factors in an information age: The promise of a living lab approach. *Proceedings of the Human Factors, Medicine, and Technology Symposium*, Baltimore, MD: The University of Maryland Medical Center.

McNeese, M. D. and M. Vidulich, Eds. (2002). *Cognitive Systems Engineering in Military Aviation Environments: Avoiding Cogminutia Fragmentosa.* Dayton, Ohio, Wright Patterson Air Force Base, CSERIAC Press.

Rasmussen, J. (1999). Ecological interface design for reliable human - machine systems. *International Journal of Aviation Psychology*, *9* (3), 203-223.

Woods, D. D. (1998). Designs are hypotheses about how artifacts shape cognition and collaboration. *Ergonomics*, *41*,168-173.

Zsambok, C. E., and G. Klein, (Eds.) (1997). *Naturalistic decision making.* Mahway, NJ: Erlbaum.